Die Rechnungen wurden an der IBM-7090 und an der IBM-360/91 in Garching durchgeführt.

Für die Unterstützung unserer Arbeiten auf dem Gebiet der Röntgenstrukturanalyse organischer Verbindungen danken wir der Deutschen Forschungsgemeinschaft, dem Fonds der Chemie und der Badischen Anilin- und Soda-Fabrik.

Literatur

AMIT, A., BRANDL, F., BRODHERR, N., GIEREN, A., HÄDICKE, E., HOPPE, W., HUBER, R. & RÖHRL, M. (1967). Vortrag auf der 9. Diskussionstagung der Sektion für Kristallkunde der Deutschen Min. Ges., 24.–27. April, Bonn und Jülich, 53–55. Bonn: Druck Keese.

COCHRAN, W. (1955). *Acta Cryst.* **8**, 473.

HECKER, E. (1967). *Naturwissenschaften*, **54**, 282.

HECKER, E., BARTSCH, H., BRESCH, H., GSCHWENDT, M., HÄRLE, E., KREIBICH, G., KUBINYI, H., SCHAIRER, H. U., v. SZCZEPANSKI, C. & THIELMANN, H. W. (1967). *Tetrahedron Letters*, **33**, 3165.

HECKER, E., BRESCH, H., GSCHWENDT, M., HÄRLE, E., KREIBICH, G., KUBINYI, H., SCHAIRER, H. U., v. SZCZEPANSKI, C. & THIELMANN, H. W. (1966). *Z. anal. Chem.* **221**, 424.

HECKER, E., HÄRLE, E., SCHAIRER, H. U., JACOBI, P., HOPPE, W., GASSMANN, J., RÖHRL, M. & ABEL, H. (1968). *Angew. Chem.* **80**, 913–914.

HECKER, E. KUBINYI, H., v. SZCZEPANSKI, C., HÄRLE, E & BRESCH, H. (1965). *Tetrahedron Letters*, **23**, 1837.

HOPPE, W. (1957). *Z. Elektrochem.* **61**, 1076.

HOPPE, W., BRANDL, I., STRELL, I., RÖHRL, M., GASSMANN, J., HECKER, E., BARTSCH, H., KREIBICH, G. & v. SZCZEPANSKI, C. (1967). *Angew. Chem.* **79**, 824.

HOPPE, W., ZECHMEISTER, K., RÖHRL, M. & BRANDL, F. (1969). *Tetrahedron Letters*, **35**, 667.

KOPFMANN, G. & HUBER, R. (1968). *Acta Cryst.* A**24**, 348.

PEERDEMAN, A. F., VAN BOMMEL, A. J. & BIJVOET, J. M. (1951). *Proc. K. Ned. Akad. Wet.* B**54**, 16.

PETTERSEN, R. C., BIRNBAUM, G. J., FERGUSON, G., ISLAM, K. M. S. & SIME, J. G. (1968). *J. Chem. Soc.* (*B*), p. 980.

PETTERSEN, R. C., FERGUSON, G., CROMBIE, L., GAMES, M. L. & POINTER, D. J. (1967). *Chem. Comm.* 716.

SAYRE, D. (1952). *Acta Cryst.* **5**, 60.

ZECHMEISTER, K. (1969). Dissertation TH München.

Acta Cryst. (1971). B**27**, 1730

Tangent Formula Applications in Protein Crystallography: An Evaluation

BY CHARLES L. COULTER

Department of Anatomy, University of Chicago, Chicago, Illinois 60637, *U.S.A.*

WITH AN APPENDIX BY R. B. K. DEWAR

Illinois Institute of Technology, Chicago, Illinois, U.S.A.

(*Received* 4 *May* 1970)

The tangent formula has been applied to X-ray data from crystals of vitamin B_{12}-5′-phosphate and myoglobin to evaluate its effectiveness as a method for structure elucidation for large molecule crystal structures. A stable, self-consistent phase solution for the tangent formula has been shown to exist for the orthorhombic B_{12}-phosphate data. This solution was near the correct solution and could be reached through partial atom phasing at 1·2, 2 or 2·5 Å, followed by tangent-formula refinement. A test calculation also indicated that a single isomorphous derivative could have been used to solve the structure. A method for calculating the standard deviation of the vector distribution of a tangent formula phase prediction has been described; this can be used as a weighting factor for Fourier series calculations. The weighted $|F|$ maps for B_{12}-phosphate were better than the E maps, with less spurious density and sharper peaks.

Calculations for myoglobin were less successful, partly because $P2_1$ is a polar low-symmetry space group. Phases were derived for 1·4 Å data by starting with the most reliable isomorphous replacement phase angles to 2 Å resolution, and holding 10% of these constant during refinement. The weighted $|F|$ maps and E maps calculated with refined phases were comparable with the 2 Å isomorphous replacement electron density maps from which the structure was derived. The tangent formula should be more useful for protein crystals of higher symmetry, as suggested by the B_{12}-phosphate experiments.

Approaches for using the tangent formula with proteins are discussed, and a description of the program is included as an Appendix.

Introduction

The tangent formula is an expression relating the phase angle of a crystallographic reflection to the phases of other reflections within the data set. It was derived by Karle & Hauptman (1956) from a consideration of the theory of the joint probability distribution of phase angles in noncentrosymmetric crystal structures. The

equation has proved to be a valuable tool for predicting probable values for phase angles based on a small number of known related phases in a variety of crystal structures (Karle, 1964; Karle & Karle, 1966*b*). Initial phases are either derived directly by applying inequalities or statistical formulas (Karle & Karle, 1966*a*) or indirectly by locating some of the atoms in the unit cell by the Patterson method or other techniques (Karle, 1968; Coulter, Hawkinson & Friedmann, 1969).

The use of direct methods in protein crystallography was first considered by Crick (cited in Cochran, 1952) for hemoglobin. Cochran (1952, 1955) analyzed several phase relations from both a statistical and a Fourier series point of view, and also explicitly discussed the question of phase extension (*i.e.* the prediction of phase angles for high $\sin\theta$ reflections from known phases for lower $\sin\theta$ reflections). Recently, primary emphasis in protein crystallography has been placed upon the use of the tangent formula and other techniques to improve or extend a set of phases based on one or more crystalline heavy-atom derivatives isomorphous with the parent protein crystals. Bijvoet (1952) has discussed the relationship of isomorphous replacement to phase determination. Coulter (1965) and Weinzierl, Eisenberg & Dickerson (1969) have explored the use of the tangent formula to choose between the two possible phase angles indicated for a general reflection by a single isomorphous derivative. Weinzierl *et al.* also investigated the improvement of low resolution phases by this means, and Reeke & Lipscomb (1969) studied both phase improvement and phase extension. Hoppe (1963), Gassmann (1966), Hoppe & Gassmann (1964, 1968) and Karle (1968) have considered approaches involving Fourier, least-squares and probability methods for improving phases for partially known or poorly determined structures. The emphasis in this paper is on applications of the tangent formula to X-ray diffraction data from large biological molecules.

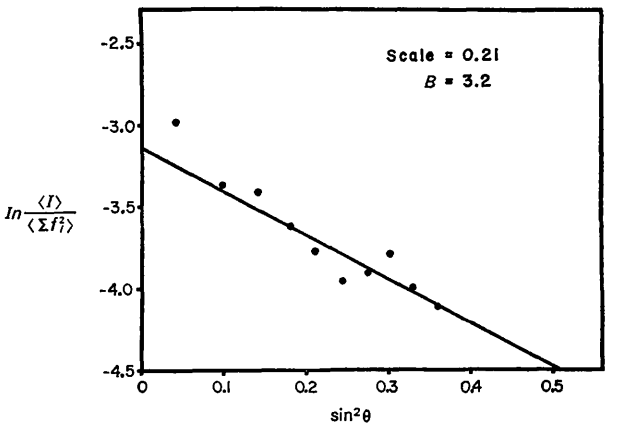

Fig. 1. Wilson plot for the vitamin B_{12}-5′-phosphate data. General reflection, $|E_{\mathbf{h}}|^2=|F_{\mathbf{h}}|^2/\sum_{j=1}^{N} f_{j\mathbf{h}}^2$.

Analysis

Hughes (1953) derived the equation:

$$U_{\mathbf{h}}=N\langle U_{\mathbf{k}}U_{\mathbf{h}-\mathbf{k}}\rangle_{\mathbf{k}}\,, \tag{1}$$

which is applicable to a crystal structure containing N equal atoms. Here, $U_{\mathbf{h}}(=F_{\mathbf{h}}\Big/\sum_{j=1}^{N} f_j)$ is the unitary structure factor for reflection $\mathbf{h}(=h,k,l)$. Cochran (1955) and Karle & Karle (1966*a*) derived similar expressions from probability considerations; the following equation is from the Karles' paper:

$$\mathscr{E}_{\mathbf{h}}=N^{1/2}\langle\mathscr{E}_{\mathbf{k}}\mathscr{E}_{\mathbf{h}-\mathbf{k}}\rangle_{\mathbf{k}}\,, \tag{2}$$

where the $\mathscr{E}_{\mathbf{h}}$ are quasi-normalized structure factors (Karle & Hauptman, 1959) defined as:

$$\mathscr{E}_{\mathbf{h}}=\sigma_2^{-1/2}\sum_{j=1}^{N} Z_j\exp(2\pi i\mathbf{k}.\mathbf{r}_j)\,. \tag{3}$$

Here Z_j is the atomic number of the jth atom having coordinates represented by the vector $\mathbf{r}_j$ in a unit cell containing N atoms, and $\sigma_n=\sum_{j=1}^{N} Z_j^n$. Equations (1) and (2) require that the expected values of the amplitudes, $\langle|\mathscr{E}_{\mathbf{h}}|\rangle$, and phases, $\langle\varphi_{\mathbf{h}}\rangle$, be related to $|\mathscr{E}_{\mathbf{k}}\mathscr{E}_{\mathbf{h}-\mathbf{k}}|$ and $(\varphi_{\mathbf{k}}+\varphi_{\mathbf{h}-\mathbf{k}})$ respectively for noncentrosymmetric systems. Karle & Hauptman (1956) have found an approximate relation, equation (4), by probability methods which does not assume equal atoms but is clearly comparable with equation (1) or (2).

$$\mathscr{E}_{\mathbf{h}}\simeq\sigma_2^{3/2}\sigma_3^{-1}\langle\mathscr{E}_{\mathbf{k}}\mathscr{E}_{\mathbf{h}-\mathbf{k}}\rangle_{\mathbf{k}}\,. \tag{4}$$

If we express the complex number $\mathscr{E}_{\mathbf{h}}$ in terms of its real and imaginary parts and divide we get equation (5)

$$\tan\varphi_{\mathbf{h}}\simeq\frac{\langle|\mathscr{E}_{\mathbf{k}}\mathscr{E}_{\mathbf{h}-\mathbf{k}}|\sin(\varphi_{\mathbf{k}}+\varphi_{\mathbf{h}-\mathbf{k}})\rangle_{\mathbf{k}}}{\langle|\mathscr{E}_{\mathbf{k}}\mathscr{E}_{\mathbf{h}-\mathbf{k}}|\cos(\varphi_{\mathbf{k}}+\varphi_{\mathbf{h}-\mathbf{k}})\rangle_{\mathbf{k}}} \tag{5}$$

which is the tangent formula. This relation was derived and applied by Karle & Hauptman (1956). Karle & Karle (1966*a*) have shown that equation (4) can be used to derive equation (6):

$$\varphi_{\mathbf{h}}\simeq\langle\varphi_{\mathbf{k}}+\varphi_{\mathbf{h}-\mathbf{k}}\rangle_{\mathbf{k}_r} \tag{6}$$

where the average is now over the subset of indices, $\mathbf{k}_r$, corresponding to large normalized structure factors. Equation (6), the $\sum_2$ relation, is analogous to equations (1) and (2), and was suggested by the analyses of Karle & Hauptman (1950) and Sayre (1952); this relation is extensively used in the early stages of the direct solution of the phase problem for noncentrosymmetric crystal structures.

Some important operational aspects of equation (6) were considered by Cochran (1955); he concluded that the $\sum_2$ relation is best satisfied when $\int_v \varrho_p^3 dV$ is a maximum positive, where ϱ_p is a partial term Fourier series with coefficients based on point atoms. Equation (6) corresponds to the unweighted tangent formula applied

to reflections with large values of E (normalized structure factors, E_h, are used in place of the quasi-normalized $\mathscr{E}_h$ values in application). Thus one view of the predicted behavior of the tangent formula is that it will adjust phases so as to increase density at points where it exists in the partial Fourier series, thus maximizing $\int_v \varrho_p^3 \mathrm{d}V$. The maximization condition is necessarily satisfied to a greater extent when the structure contains a region of high electron density. Phase refinement can thus, in the extreme case, lead to a single heavy atom in the unit cell. This solution fits equation (6) very well, and equation (5) less well since the calculated amplitudes would not agree with $|E|_{obs}$. Cochran (1952) has shown that the probabilitiy of ϱ_p having significant positive values at atom sites is high even with very few terms in the Fourier series. Thus the normal result on refinement is to increase density at atom sites, revealing the structure. Cochran's criterion for fit is a useful general guide to monitoring tangent-formula behavior. It also appears to be compatible with the observation of Reeke & Lipscomb (1969) that refined phase angles for the carboxypeptidase-A data gave sharper Fourier peak definition accompanied by distortion of some groups.

Experimental

A number of different techniques for applying the tangent formula have been used. The most obvious approach is to take all the data for which $|E| \geq 1{\cdot}5$ or $1{\cdot}0$ and for which phase angles are available and iteratively refine these data to convergence. Data to be used in predicting phases [*i.e.* on the right side of equation (5)] are selected on the basis of previous phase shifts, the magnitude of the calculated E (Karle & Karle, 1966*b*) or other criteria. This procedure generally gives acceptable results with structures for which full data are available when the initial phases are assigned on the basis of part of the structure, located in the cell by other means. This procedure is less controlled and more prone to give trivial solutions or to diverge than the other methods. A more sophisticated approach utilizes the most accurate phase angles initially, and only gradually includes other large E phases in the predicting group. The initial assessment of accuracy can be made by Sim's (1959, 1960) method based on the distribution of phases, by $|F_o|-|F_c|$ agreement (Karle, 1968) or by using figures-of-merit from protein multiple isomorphous replacement phase-determination (Dickerson, Kendrew & Strandberg, 1961). A variant of this method is to hold the initial phases, judged to be the most accurate, constant and to use these data to predict phase angles for other reflections having large values of E. The most reliable of the newly predicted phases are then added to the initial group for further prediction cycles. The latter two procedures have been successfully applied to biological crystal structures. The final solution is the same as that reached by the first technique for cases where both methods have been applied, provided all phases are eventually allowed to vary.

X-ray diffraction data from crystals of vitamin B_{12}-5′-phosphate (Coulter, Hawkinson & Friedmann, 1969; Hawkinson, Coulter & Greaves, 1970) and myoglobin (Bodo, Dintzis, Kendrew & Wyckoff, 1959; Kendrew, 1963) were used in this study. The B_{12}-phosphate data were collected from an air dried crystal using a PAILRED linear diffractometer and Cu $K\alpha$ radiation from a silicon monochromator. All reflections to $2\theta=75°$ ($d_{min}=1{\cdot}2$ Å) were scanned and 2112 observed reflections were used in the analysis. The myoglobin data to 2 Å resolution were recorded on precession films which were then scanned using Joyce–Loebl recording microdensitometers. The 2 to 1·4 Å data were collected with a linear diffractometer (Arndt & Phillips, 1961) and considerably overlapped the 2 Å data set so as to allow for scaling the two sets of intensities together. There were 9250 observed reflections to 2Å, and 16,858 observations in the 1·4 Å data set. Very few reflections were observed beyond 1·3 Å for vitamin B_{12}-5′-phosphate, and some high resolution myoglobin data extended to 1·3 Å; thus the data for these two cases were of comparable resolution. The B_{12}-phosphate crystals are orthorhombic, space group $P2_12_12_1$, with $a=23{\cdot}72$, $b=21{\cdot}74$, $c=16{\cdot}07$ Å and four B_{12}-phosphate molecules and about 60 water molecules per unit cell. Spermwhale myoglobin crystals are monoclinic, space group $P2_1$, with $a=64{\cdot}5$, $b=30{\cdot}86$, $c=34{\cdot}7$ Å, $\beta=106°$ and 2 protein molecules per cell, along with appreciable amounts of solution [aqueous $(NH_4)_2SO_4$]. The experiments with the vitamin B_{12}-5′-phosphate data were more extensive than those with myoglobin because the structure is small enough to permit extensive analysis under different conditions; also, the orthorhombic symmetry makes it a more favorable case for tangent-formula applications. The two crystal structures are not directly comparable, but they are similar enough in resolution of the data and moleculer features for the B_{12}-phosphate results to suggest what might be done

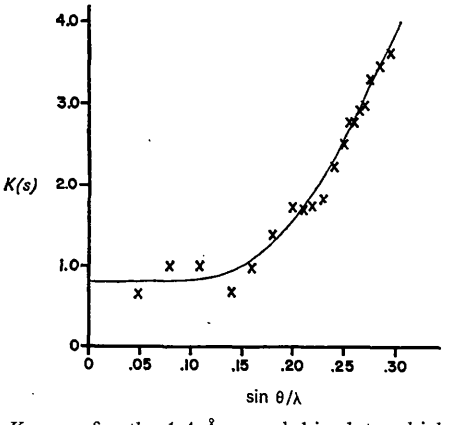

Fig. 2. K curve for the 1·4 Å myoglobin data which was used to scale and correct the data for temperature factor effects (Karle & Hauptman, 1953).

with protein data from crystals of orthorhombic or higher symmetry.

In order to apply the tangent formula to these structures, the amplitudes first had to be normalized to give $|E|$ values. For vitamin B_{12}-5′-phosphate the temperature factor and scale factor were derived from the Wilson (1942) plot shown in Fig. 1. For myoglobin, the *K*-curve method suggested by Karle & Hauptman (1953) was used, since this does not assume an exponential intensity decrease with $\sin\theta$. The results for the 2 Å myoglobin data are summarized in Table 1, and Fig. 2 shows the correction curve for the myoglobin data.

Table 1. *Myoglobin E statistics*

	Observed		Theoretical	
Quantity	*h0l*	*hkl*	Centric	Acentric
$\langle\|E\|\rangle$	0·857	0·897	0·798	0·886
$\langle\|E\|^2\rangle$	1·11	1·02	1·000	1·000
$\langle\|E\|^2-1\rangle$	1·02	0·742	0·968	0·736
% > 1·0		37	32	37

Tangent formula experiments

The experiments in which the B_{12}-phosphate data were used are summarized in Table 2. Three to 5 cycles of tangent-formula refinement were carried out in each of these experiments using program *TANG* described in the Appendix. All the phases were predicted in each cycle, and reflections for which the predicted phase shift was over 50° or for which $|E|_{calc}$ was less than 0·5 were not used in predicting other phases. Phase-angle comparisons and Fourier syntheses were used to evaluate the results. It was somewhat evident in the B_{12}-phosphate work, and very evident in the early protein studies, that *E* maps were not very satisfactory substitutes for electron density maps in cases where the resolution of the data was poor and the structures were not highly over-determined. In addition, for the protein case, the tangent formula seems likely to be most useful in conjunction with isomorphous replacement phase information. One alternative to an *E* map is a weighted Fourier series calculation, with the coefficients assigned a weight based on the phase-angle reliability as suggested by Blow & Crick (1959). In the multiple isomorphous replacement method the figure-of-merit (Dickerson, Kendrew & Strandberg, 1961) corresponds roughly to the cosine of the mean error in phase angle for a reflection, and is used as a weighting factor in the Fourier synthesis. Karle & Karle (1966*a*) have derived a variance formula based on the phase-probability distribution which is useful in the early stages of the direct determination of phase angles using the tangent formula. Variance estimates for cytochrome-c (Weinzierl *et al.*, 1969) and carboxypeptidase-A (Reeke & Lipscomb, 1969) were only weakly correlated with the correctness of the phases, and the standard deviation of the vector distribution itself may be more useful. Each contributor to equation (5) for a given phase prediction is a vector of magnitude $|E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}|$ and phase $\varphi_{\mathbf{k}}+\varphi_{\mathbf{h}-\mathbf{k}}$; the sharpness of the total vector distribution is reflected in the fraction of the maximum possible contribution found for this phase, which is *r* in Fig. 3.

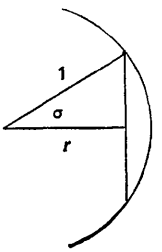

Fig. 3. Fraction of maximum possible contribution found for phase $\varphi_{\mathbf{h}}$.

Expressing equation (5) as $\tan\varphi_{\mathbf{h}}=N/D$, *r* is given by equation

$$r=(N^2+D^2)^{1/2}\Big/\sum_{\mathbf{k}}|E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}|\,. \qquad (7)$$

The angle σ [not related to the σ's of equations (3) and (4)] is then defined relative to the chord normal to *r*, and

$$\tan\sigma=(1-r^2)^{1/2}/r\,. \qquad (8)$$

Table 2. *Tangent formula calculations*

Experiments 1–6 performed with B_{12}-phosphate data, experiments 7 & 8 with myoglobin data

Experiment	Å	Phasing	Starting set No. of reflections	$\|E\|_{min}$	Phases out*	$\|E\|_{min}$	*R*%†	$\Delta\alpha$
1	1·2	5 atoms	673	1·0	676	0·93	24·2	(0)
2	1·2	100 atoms	673	1·0	783	0·9	25·4	3°
3	2·0	5 atoms	230	0·9	688	0·9	28·4	3°
4	2·5	5 atoms	129	0·9	348	0·9	23·4	9°
5‡	1·2	5 atoms	733	0·9	640	0·9	26·7	–
6‡	1·2	Isomorphous§	355	1·0	930	0·8	28·6	–
7	2·0	Calc	1546	1·5	1017	1·5	26·6	–
8	2·0	Isomorphous	1546′	1·5	3191	1·2	29·2	–

* Phase angles were predicted to 1·2 Å (1–6) or 1·4 Å (7, 8) resolution in each case.
† *R* values is calculated after scaling $|E|_{calc}$ to $|E|_{obs}$.
‡ Experiments based on calculated parent amplitudes.
§ Phases based on calculated parent and observed derivative amplitudes (see text).
′ 154 of these phase angles were held constant (see text).

Simaika (1956) has studied circular distributions such as this, and showed that angle σ is the standard deviation of the resultant vector distribution. Fig. 4 illustrates an empirical confirmation of this for the B_{12}-phosphate data. This standard deviation can now be used as a Fourier weighting factor, $\cos \sigma_h$, for coefficient $|F|_h$ and the resultant Fourier synthesis should give an electron density map with minimum mean-square error averaged over the cell (Blow & Crick, 1959). Additional coefficients can be included with isomorphous replacement phase angles and figure-of-merit weights. Another alternative would be to treat the refined phase set as though it were another derivative. The calculated standard deviations follow $|E|_{calc}$, as expected, since the magnitude of r is itself an $|E|_{calc}$ for a unit-radius phase circle. In practice, using the σ's in weighting electron density map coefficients allows the tangent formula to be applied to reflections with lower $|E|$ values, thereby increasing the number of Fourier coefficients. Reeke & Lipscomb (1969) have shown that good quality, high-resolution protein maps can be calculated using only the largest reflections, so neglecting the very low E data should not be a problem.

The crystal structure of vitamin B_{12}-5′-phosphate was solved by locating the cobalt atom and the 3′-PO_3 group from a sharpened Patterson summation, refining the phase angles based on these 5 atomic positions with the tangent formula, and examining an E map calculated with the refined phases. Positions for about 75% of the atoms in the molecule could be assigned from the E map, and the complete structure was derived and refined using Fourier and least-squares methods (Hawkinson, Coulter & Greaves, 1970). Initial phase angles for experiments 1, 3, 4 and 5 in Table 2 were based on these 5 atomic positions. As a control, 5 other atomic positions, with the cobalt in common, were used to calculate initial phases in one case. These phases refined to the same solution ($\pm 3°$) observed in experiment 1. The single isomorphous derivative studies, experiments 5 and 6, were done using calculated parent amplitudes based on the atomic positions for the vitamin B_{12}-5′-phosphate molecule, excluding the 5′-phosphorus atom and the 2 phosphate oxygen atoms which were not replaced by water molecules in the air-dried vitamin B_{12} crystal structure (Hawkinson, Coulter & Greaves, 1970). The observed B_{12}-phosphate intensities were used as the heavy-atom derivative data. Vitamin B_{12} and vitamin B_{12}-5′-phosphate are crystallographically very similar, but not sufficiently isomorphous to permit isomorphous replacement phasing using the observed intensities for both compounds.

Myoglobin data were used for experiments 7 and 8, Table 2. The calculated phase angles in experiment 7 were based on positional parameters supplied by Kendrew and Watson. There were 1546 reflections in the 2 Å data set with $|E| \geq 1{\cdot}5$; isomorphous replacement phase angles and figures-of-merit had been derived for these data (Kendrew, Dickerson, Strandberg, Hart, Davies, Phillips & Shore, 1960) and were used in experiment 8 and in phasing isomorphous replacement Fourier sections for comparison with the same sections calculated using refined phase angles.

Results and discussion

Complete Fourier synthesis maps were examined in all of the B_{12} experiments. Fig. 5 shows a representative section from the final electron density map for vitamin B_{12}-5′-phosphate (*a*) along with the same section from the E map (experiment 1) shown in (*c*) and from a weighted Fourier synthesis calculated using 973 coefficients with $|E| \geq 0{\cdot}5$ illustrated in (*b*). The E map was used to solve the structure, but the weighted $|F|$ map, calculated later, had better defined peaks and less spurious density. For example, carbon atom R3 of the ribose [C(3′)] appeared at maximum density on the $x=46/60$ section of the weighted map, very near the final x coordinate of 46·6/60; on the E map this peak was spread over about 1·5 Å in x, and had significant density up to $x=50/60$. The spurious E map peak marked with the arrow in Fig. 5 was much lower in the weighted $|F|$ synthesis, and other spurious density had decreased. The refined phases in experiment 2, based on 100 atomic positions, were the same as those based initially upon 5 positions. This experiment was done in an effort to objectively place solvent molecules in the cell. The solvent region was not clear in either of the E maps calculated with the refined phase angles, and was only slightly better in the weighted $|F|$ syntheses. However, these experiments did show that a stable self-consistent solution to the tangent formula existed, and that this solution was very close to the correct phase solution derived by standard structure analysis and refinement techniques.

The effect of resolution on the behavior and reliability of the tangent formula was examined in experiments 3 and 4. Phase extension is an area where Hoppe & Gassmann (1964, 1968) and co-workers have been especially active and to which Reeke & Lipscomb

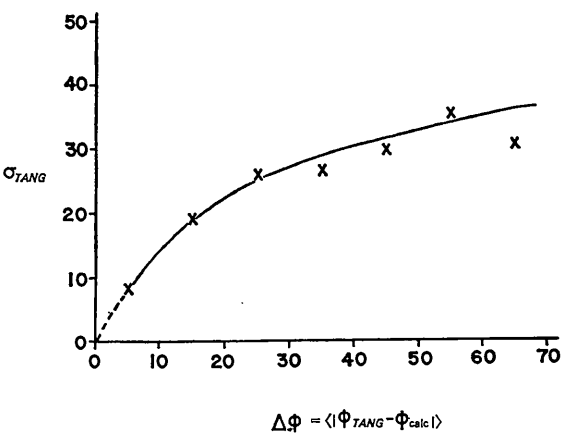

Fig. 4. Calculated standard deviation *versus* phase error for B_{12}-phosphate data.

(1969) and Barrett & Zwick (1970) have recently contributed some suggestions. These experiments are similar to those of Reeke & Lipscomb (1969) as they involve using only the tangent formula to extend phase predictions to higher resolutions. The results of the 1·2 Å phase predictions beginning from 2 and 2·5 Å data are summarized in Table 3. The 2 Å extended refinement has converged to the original E map phase solution, and the 2·5 Å refinement solution is similar, but a bit further away. The initial number of phases in the 2·5 Å experiment was almost too small, and these data did not interact well with the higher resolution reflections in many cases as can be seen from the smaller number of phase angles which were well determined by the $|E|_{calc}$ criterion. Some E map sections for these two experiments are given in Fig. 6, and can be compared with the original E map section in Fig. 5.

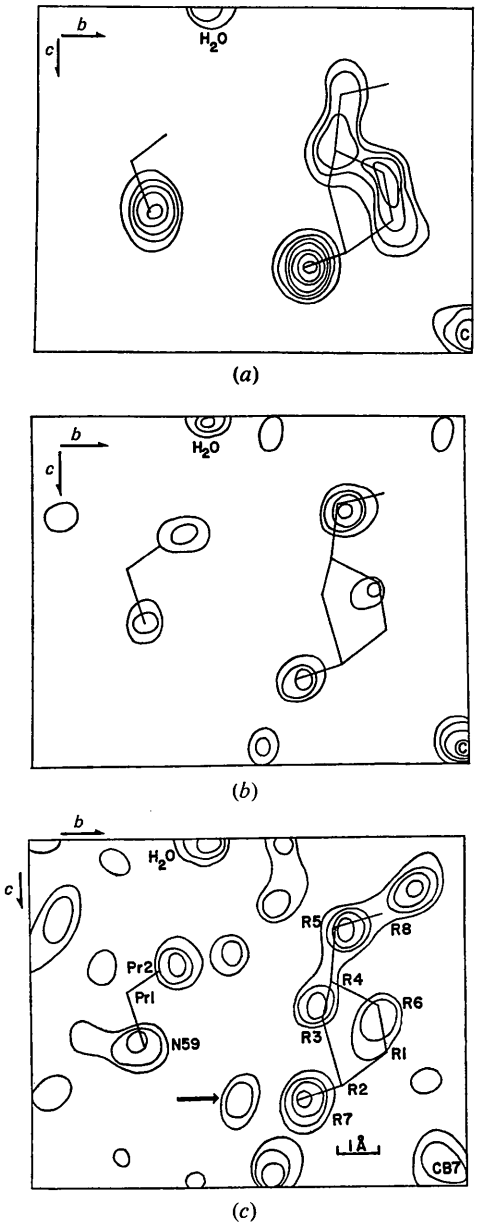

Fig. 5. Fourier sections for the B_{12}-phosphate structure: (*a*) final observed Fourier synthesis ($\varrho=10$); (*b*) weighted $|F|$ map and (*c*) E map. Coordinate $x=50/60$.

Table 3. *Comparison of refined phase angles for vitamin* B_{12}*-5′-phosphate based on different initial resolution data. Average agreement for these cases for all the data is given in Table* 2

hkl	$\|E\|$	Refined phase*	Refined phase†	Refined phase‡
352	2·16	239°	237°	247°
921	1·91	177	175	174
216	1·80	245	240	260
291	1·78	88	86	81
253	1·77	263	257	255
511	1·74	5	9	16
431	1·66	108	119	109
831	1·66	161	165	142
432	1·52	162	155	151
456	1·43	33	34	55
274	1·24	283	292	282
653	1·23	243	231	229
424	1·19	101	92	88
915	0·92	103	89	84

* Derived from starting phase set of 1·2 Å resolution (experiment 1)

† Derived from starting phase set of 2 Å resolution (experiment 3)

‡ Derived from starting phase set of 2·5 Å resolution (experiment 4)

Coulter (1965) suggested using the tangent formula in conjunction with a single highly isomorphous derivative to derive a set of phase angles for a protein crystal structure. This possibility could not be tested with the myoglobin data because the y coordinates of the substituents in the double derivative for which data are available are nearly the same (Bodo *et al.*, 1959). Neither this derivative nor the single-site derivatives can be easily used for crystals of space group $P2_1$ since the predicted phases are centered about α or $\alpha+\pi$ for a given k index. Karle (1966) has suggested a method of phase determination for the case where the isomorphous replacement contribution is always real, but this method was not tested here. As noted before, vitamin B_{12}-phosphate is not sufficiently isomorphous with vitamin B_{12} to be used as a derivative, so experiments 5 and 6 were done with parent amplitudes calculated from the B_{12}-5′-phosphate positional parameters omitting the 5′-PO_2 atoms. These exact parent amplitudes were used with the observed B_{12}-phosphate derivative amplitudes and the heavy-atom structure factor was calculated from the positional parameters for the additional 3 atoms in the derivative. Two phase angles were possible for each general reflection in the parent data set. The 355 starting reflections used in the tangent formula had the two predicted phases within 150° of each other. Fourier maps from experiments 5 and 6 were virtually identical, and were very similar to the E map

from experiment 1. Apparently this orthorhombic structure could have been solved with a single isomorphous PO_2 derivative.*

Myoglobin crystallizes in a polar monoclinic space group, $P2_1$, and represents a more difficult case for tangent formula applications than vitamin B_{12}-5′-phosphate. Karle (personal communication) and others have noted that incorrect self-consistent phase sets fulfilling the Σ_2 relation [equation (6)] are often found in polar space groups. These solutions usually involve a pseudo-mirror plane normal to the polar direction. When the molecule contains a relatively heavy atom, the tendency to converge to this solution or a similar one involving very high density at the heavy-atom site is more pronounced since the true phase angles reflect the heavy-atom position. Some of the results reported by Weinzierl *et al.* (1969) on cytochrome-c data (space group $P4_1$) can be understood in this way, and the phenomenon has also been noted by Reeke & Lipscomb (1969), in small structures, and for myoglobin. For small structures the difficulty has been surmounted in most cases by holding the initial phase angles which are considered most reliable constant during tangent formula applications and only relaxing this constraint (if at all) when most of the data have been included. This same procedure has been used here with the myoglobin data, and was used by Reeke & Lipscomb (1969) with carboxypeptidase A–(also in space group $P2_1$).

Experiment 7, Table 2 was performed with calculated phases to establish the existence of a stable phase solution. It was evident that additional reflections beyond the original set had to be added very slowly to avoid convergence to a solution with a very large iron-atom peak. If the initial data set was refined to convergence (2–3 cycles) and these phases were then fixed, other phase angles to a resolution of 1·4 Å could be determined and included in the further predictions. This procedure led to a stable solution comparable to the true solution as judged by the similarity of the E map to the isomorphous replacement Fourier synthesis at 2 Å. Experiment 8 was performed, using the less accurate isomorphous replacement phase angles. Successful refinement was attained by beginning with the data for which $|E| \geq 1{\cdot}5$ and the figure-of-merit exceeded 0·9 (*i.e.* the most reliable 10% of the isomorphous phase angles) and holding these phase angles constant throughout the refinement. The refinement then proceeded in stages, with 3 cycles on the $|E| \geq 1{\cdot}5$ data followed by 2 cycles on reflections down to $|E| \geq 1{\cdot}2$. After each cycle the most reliable new phase angles (large $|E|_{calc}$ and small shift in angle) were included in the next prediction cycle. Convergence behavior is illustrated in Fig. 7, and a graph of $|\varphi_{refined} - \varphi_{isom}|$ *versus* the figure of merit is given in Fig. 8. Sections from a weighted Fourier synthesis and an $|E|$ map, both calculated using 3191 coefficients with $|E| \geq 1{\cdot}2$ and $d \geq 1{\cdot}4$ Å, are shown in Fig. 9 along with the corresponding sections from the 2 Å isomorphous replacement Fourier used to derive the structure.

In the sections thus far examined, the weighted Fourier map is quite comparable in quality to the original myoglobin $|F|$ best Fourier synthesis, and the $|E|$ map is of similar quality except near the iron atom. Since the phases initially used correspond to about 10% of the data, this suggests that a single isomorphous replacement derivative, which would also accurately define about 10% of the large $|E|$ phase angles, could possibly have been used to solve the structure had an appropriate derivative been available. The result also demonstrates that the tangent formula can be used with caution to improve isomorphous replacement phase angles that are poorly determined. A similar conclusion for low-resolution phases was reported by Reeke & Lipscomb (1969). When the best of the beginning phase angles were not held constant, the system was only stable for 2 to 3 cycles after which it diverged or converged to a trivial solution; this behavior was also noted in the earlier experiments with isomorphous phase angles (Weinzierl *et al.*, 1969; Reeke & Lipscomb, 1969). In view of the problems connected with preparing an appropriate isomorphous replacement derivative in this space group, and the rather less satisfactory behavior of the tangent formula because of the low

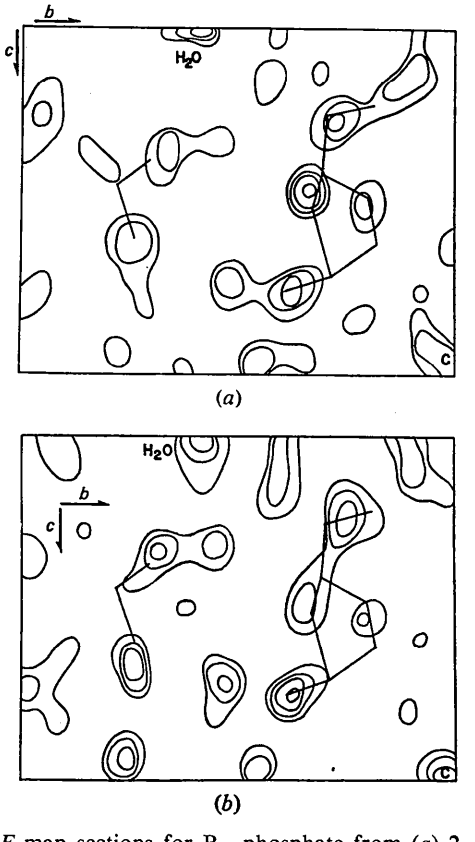

Fig. 6. E map sections for B_{12}-phosphate from (*a*) 2 Å and (*b*) 2·5 Å starting phases. Coordinate $x = 50/60$.

* Problems with heavy atom ghost peaks can arise in some cases, especially for heavy metal substituets.

multiplicity and polar nature of the space group, it seems unlikely that the single isomorphous replacement method of solution is the best approach to solving the myoglobin structure. Phase extension was possible, but it was also affected by the above considerations. The iron-atom peak in every case tended to be large (2 to 3 times the expected height relative to other peaks) and the desired solution was not nearly as stable as in the B_{12}-phosphate case. The high resolution experiments with the carboxypeptidase-A data (Reeke & Lipscomb, 1969) were similar. For phase extension, Hoppe & Gassmann's suggestion (1968) to subtract the heavy-atom contribution from the observed intensities might help considerably, but this requires accurate values for the parent-derivative scale factor and heavy-atom occupancy parameters. Barrett & Zwick's approach (1970) uses a fast Fourier algorithm to permit refinement *via* Fourier transformation to real space, modification of the real space map by eliminating negative density and possibly controlling the height of the heavy-atom peak, and transformation back to reciprocal space at higher resolution. This method also provides greater flexibility in handling heavy-atom problems than the tangent formula. The $E(000)$ term is not used in the tangent formula applications, even though it will be quite large for a protein. The reason is that it simply acts as a shift-damping factor, since the assigned phase angle for the $\varphi_h+\varphi_o$ pair must be φ_h. The graphs and Fourier sections presented earlier suggest that the absence of the $E(000)$ term does not seriously affect the solution for B_{12}-phosphate or for myoglobin. To complete this evaluation, data for a protein crystal of orthorhombic or higher symmetry crystallizing in a nonpolar space group should be used. Such data were not available, and the completion of the study must await this application, preferably on an unknown structure.

I wish to thank Drs Jerome Karle, Robert Dewar, Aaron Klug, Christopher Bingham and Stuart Hawkinson for helpful discussions and assistance at various stages of this work. I thank Mrs Marsha Greaves for the figures and Drs Kendrew and Watson for the myoglobin data. This research was supported by the National Science Foundation (GB-5504 and 8103).

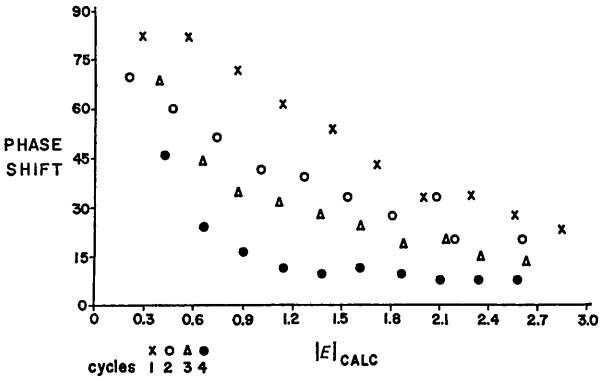

Fig. 7. Phase shift *versus* $|E|_{calc}$ for several cycles of refinement of myoglobin phases. Phase extension 2–1·4 Å.

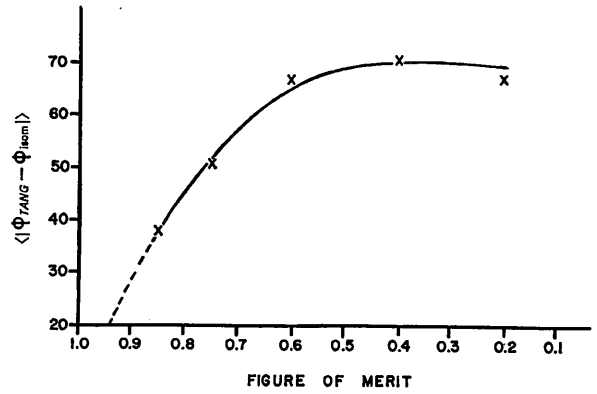

Fig. 8. Relation between tangent formula phases and isomorphous phases for myoglobin.

APPENDIX

The tangent formula program, *TANG*, is a general *FORTRAN*-IV computer program applicable to all space groups and suitable for a wide range of calculations. Fig. 10 presents a schematic summary of the operations. The needed data comprise a number of unique crystallographic reflections, some of which have phase angles assigned and the equivalent positions of the space group in Hollerith format. The program generates equivalent reflections with appropriate phase angles and stores half the sphere for these data. Reflection indices with E values and phases when known are entered on cards or tape for initial input with a key labeling each as a known fixed-phase reflection, a known variable-phase reflection or an unknown reflection with or without a phase angle. Only the known reflections are included in phase predictions with the tangent formula. The data sets are usually stored on an auxiliary tape after the first iteration. This tape has the facility for storing multiple data sets on a single file, and the past history of each reflection is also retained; this system permits editing of the data between cycles. In each iteration *TANG* calculates the phase angles of all nonfixed reflections using the two forms of the tangent formula:

$$\tan\varphi_h \approx \frac{\langle E_k E_{h-k}\sin(\varphi_k+\varphi_{h-k})\rangle_k}{\langle E_k E_{h-k}\cos(\varphi_k+\varphi_{h-k})\rangle_k} \qquad \text{(A1)}$$

$$\tan\varphi_h \approx \frac{\sum_k E_k E_{h-k}\sin(\varphi_k+\varphi_{h-k})}{\sum_k E_k E_{h-k}\cos(\varphi_k+\varphi_{h-k})} \qquad \text{(A2)}$$

The phase angles from equations (A1) and (A2) are the same, but the numerator and denominator of each equation are separately scaled to $|E|_{obs}$ over the entire data set by a least-squares method and the two resulting $|E|_{calc}$ values differ. Both are written out. The standard

deviation of the phase angle is also calculated, as noted earlier. The phase-distribution diagram can be displayed and graphs like that of Fig. 7 are printed after each cycle. These are measures of convergence since graphs of phase shift *versus* $|E|_{\text{calc}}$ or the standard deviation are sensitive to the goodness of fit. The discrepancy index between $|E|_{\text{obs}}$ and $|E|_{\text{calc}}$ is also a useful measure of convergence (but not correctness). The inclusion of predicted phases in the next cycle as 'knowns' is mainly controlled by phase shift and $|E|_{\text{calc}}$ filters.

The calculation of the tangent formula can be time consuming. At best, a quadratic dependence on the number of data items can be achieved since the total number of interactions varies as N^2. Some care is required to attain this level; simple algorithms will tend towards $N^2 \log_2 N$ or N^3 dependence. The method used in *TANG* is linear for each fixed **k** to be predicted and thus quadratic over the entire set. The data are first sorted by index with a fast sorting program based on the method of Floyd (1964). The search for pairs then proceeds in two steps. First, all h are found such that $h_{\mathbf{k}}=h_{\mathbf{h}}+h_{\mathbf{k}-\mathbf{h}}$. The required pairs nest about the point $h_{\mathbf{k}/2}$, as shown in Fig. 11.

The midpoint is located with a binary search, and two pointers A and B are moved out in a single linear sweep which detects all such pairs. The second step is to find all h such that $h_{\mathbf{k}}=h_{\mathbf{h}+\mathbf{k}}+(-h_{\mathbf{h}})$.

This allows for interactions involving the other half

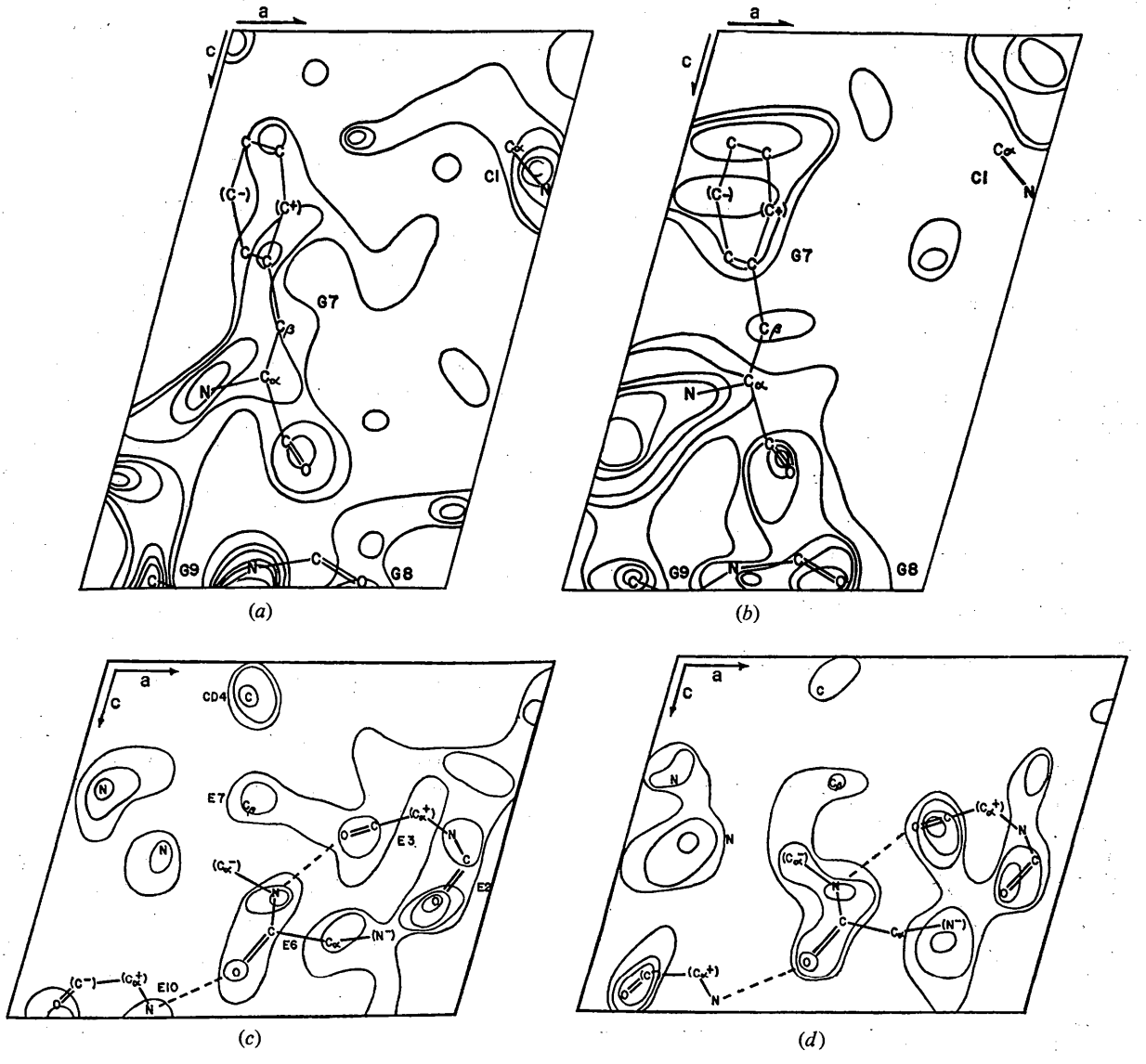

Fig. 9. Sections through the myoglobin isomorphous replacement Fourier and the Fourier maps based on refined phase angles. (*a*) α_{isom} and (*b*) weighted Fourier, are sections through phenylalanine G–7 (y coordinate=26/48); (*c*) E map (α_{TANG}) and (*d*) isomorphous replacement Fourier, cut through a portion of the E helix (y coordinate=44/48). Residue numbers are indicated, and atoms above or below these sections are labelled (+) and (−).

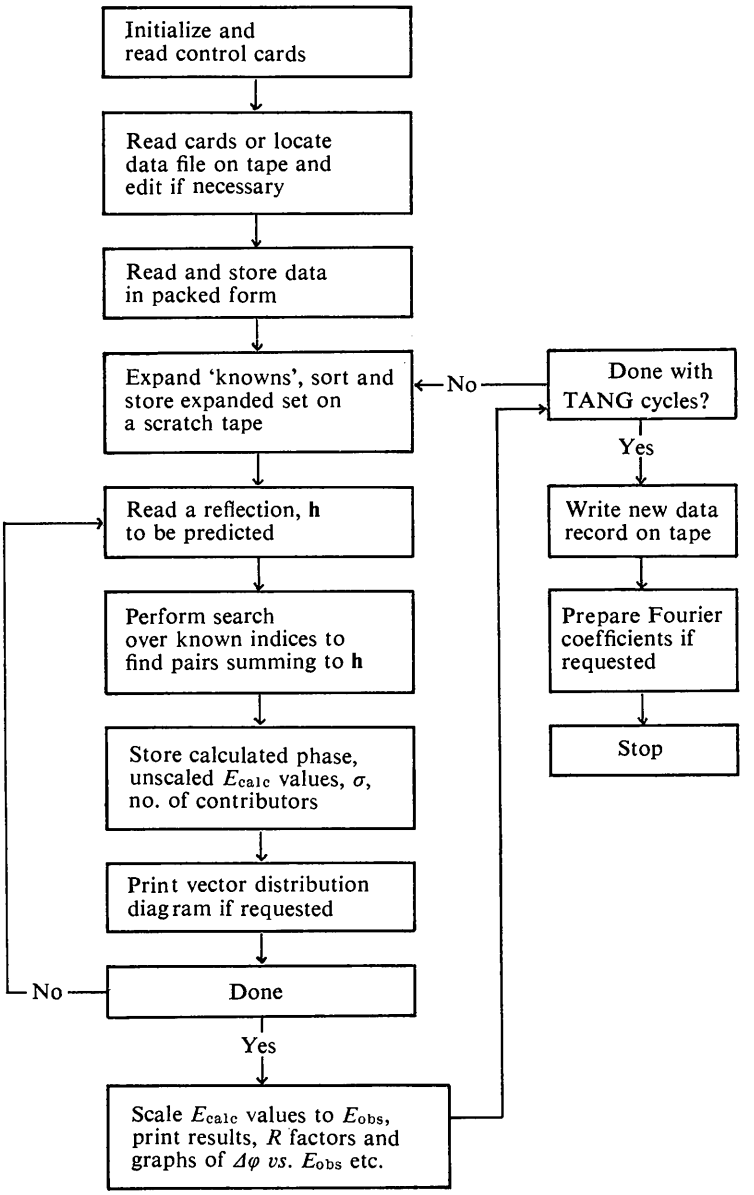

Fig. 10. Schematic flow chart for *TANG*.

of the data sphere. The phase relation in this case is the Friedel one, $\varphi_{\mathbf{h}} = -\varphi_{-\mathbf{h}}$. Fig. 12 illustrates this search. The first potential pair is (h_{f+k}, h_f) where h_f is the first entry in the list. Again a linear sweep of two pointers locates all pairs. This algorithm has been found to be efficient and straightforward. It may also be applied to the Σ_2 formula for the symbolic-addition method.

The time required for the *TANG* cycles with the vitamin B_{12}-5′-phosphate data was 3 to 8 minutes per cycle using one IBM 7094/7040 system. The longest

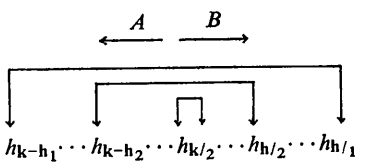

Fig. 11. First step in location of pairs (h_h, h_{k-h})

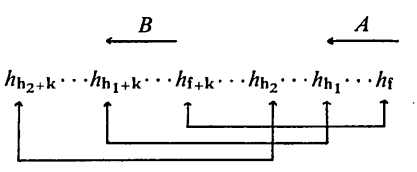

Fig. 12. Second step in location of pairs (h_h, h_{h-k}).

myoglobin run required 20 minutes to predict 6000 phases using 2000 known reflections.

References

ARNDT, U. W. & PHILLIPS, D. C. (1961). *Acta Cryst.* **14**, 807.
BARRETT, A. N. & ZWICK, M. (1971). *Acta Cryst.* A**27**, 6.
BIJVOET, J. M. (1952). Conference on *Computing Methods and the Phase Problem in X-ray Crystal Analysis:* Penn. State College, U.S.A. p. 84.
BLOW, D. M. & CRICK, F. H. C. (1959). *Acta Cryst.* **12**, 794.
BODO, G., DINTZIS, H. M., KENDREW, J. C. & WYCKOFF, H. W. (1959). *Proc. Roy. Soc.* A**253**, 70.
COCHRAN, W. (1952). *Acta Cryst.* **5**, 65.
COCHRAN, W. (1955). *Acta Cryst.* **8**, 473.
COULTER, C. L. (1965). *J. Mol. Biol.* **12**, 292.
COULTER C. L., HAWKINSON, S. W. & FRIEDMANN, H. C. (1969). *Biochim. Biophys. Acta*, **177**, 293.
DICKERSON, R. E., KENDREW, J. C. & STRANDBERG, B. E. (1961). *Acta Cryst.* **14**, 1188.
FLOYD, R. (1964). *Commun. Assoc. Comp. Mach.* **7**, 701.
GASSMANN, J. (1966). *Acta Cryst.* **21**, A6.
HAWKINSON, S. W., COULTER, C. L. & GREAVES, M. L. (1970). *Proc. Roy. Soc.* A**318**, 143.
HOPPE, W. (1963). *Z. Kristallogr.* **118**, 121.
HOPPE, W. & GASSMANN, J. (1964). *Berichte der Bunsengesellshaft.* **68**, 808.
HOPPE, W. & GASSMANN, J. (1968). *Acta Cryst.* B**24**, 97.
HUGHES, E. W. (1953). *Acta Cryst.* **6**, 871.
KARLE, J. (1964). In *Advanc. Struct. Res. by Diffn. Methods*, Vol. I, Ed. R. BRILL, p. 55. New York: Interscience.
KARLE, J. (1966). *Acta Cryst.* **21**, 273.
KARLE, J. (1968). *Acta Cryst.* B**24**, 182.
KARLE, J. & HAUPTMAN, H. (1950). *Acta Cryst.* **3**, 181.
KARLE, J. & HAUPTMAN, H. (1953). *Acta Cryst.* **6**, 473.
KARLE, J. & HAUPTMAN, H. (1956). *Acta Cryst.* **9**, 635.
KARLE, J. & HAUPTMAN, H. (1959). *Acta Cryst.* **12**, 404.
KARLE, J. & KARLE, I. L. (1966*a*). *Acta Cryst.* **21**, 849.
KARLE, I. L. & KARLE, J. (1966*b*). *Acta Cryst.* **21**, 860.
KENDREW, J. C. (1963). In *Les Prix Nobel* 1962, p. 103. Amsterdam: Elsevier.
KENDREW, J. C., DICKERSON, R. E., STRANDBERG, B. E., HART, R. G., DAVIES, D. R., PHILLIPS, D. C. & SHORE, V. C. (1960). *Nature, Lond.* **185**, 422.
REEKE, G. N. & LIPSCOMB, W. N. (1969). *Acta Cryst.* B**25**, 2614.
SAYRE, D. (1952). *Acta Cryst.* **5**, 60.
SIM, G. A. (1959). *Acta Cryst.* **12**, 813.
SIM, G. A. (1960). *Acta Cryst.* **13**, 511.
SIMAIKA, J. B. (1956). *Proc. Math. Phys. Soc. Egypt*, **5**, 113.
WEINZIERL, J. E., EISENBERG, D. & DICKERSON, R. E. (1969). *Acta Cryst.* B**25**, 380.
WILSON, A. J. C. (1942). *Nature, Lond.* **150**, 152.

Acta Cryst. (1971). B**27**, 1740

The Structure of β-Uranium

BY JERRY DONOHUE AND HOWARD EINSPAHR*

Department of Chemistry and Laboratory for Research on the Structure of Matter, University of Pennsylvania, Philadelphia, Pensylvania 19104, *U.S.A.*

(*Received* 10 *August* 1970)

During a review of the various structures of uranium, one of us discovered indexing errors that affected a number of the investigations of the β allotrope. These errors have been corrected, and least-squares refinements performed in the three possible space groups, $P4_2/mnm$, $P4_2nm$, and $P\bar{4}n2$, with resulting R_I values of 0·28, 0·24, and 0·28, respectively. It is concluded that the β-uranium powder data (Thewlis, 1952) cannot be used to determine the correct space group.

β-Uranium, the allotrope stable between 661° and 772°C, has been subject of repeated structural investigations for over 20 years. Great experimental difficulties are encountered in obtaining intensity data, owing to the excessive reactivity of the metal and to the inability of the allotrope to retain its structure on quenching.

Tucker (1951), using Cu $K\alpha$ radiation, obtained projection data at room temperature from a single crystal of a 1·4 atomic % Cr alloy of uranium that had been quenched from above the transition temperature with no evidence of reformation of the α allotrope. The lattice was found to be tetragonal with $a=10{\cdot}52$, $c=5{\cdot}57$ Å, and $Z=30$. Systematic absences ($0kl$, $k+l=2n+1$) admitted three possible space groups: the centrosymmetric $P4_2/mnm$, and the noncentrosymmetric $P4_2nm$ and $P\bar{4}n2$. A structure was proposed in $P4_2nm$ based on Patterson projections using $0kl$ and $hk0$ data. No refinement was reported. $P4_2/mnm$ was judged less favorable on the basis of detailed intensity considerations; $P\bar{4}n2$ was not considered.

Thewlis (1952) published a powder pattern of β-

* Present address: A. A. Noyes Laboratories of Chemical Physics, California Institute of Technology, Pasadena, California 91109, U.S.A.